Statistical Methods for a Tag-Recapture Experiment on a Population Closed to Recruitment: Preliminary Report

D. S. Robson

BU-374-M June, 1971

ABSTRACT

Survival rates in an animal population or subpopulation which is closed to recruitment but open to mortality may be estimated from a k-sample tag-recapture experiment. At time t_i a batch of m_i animals are captured, the tag numbers of all recaptures are recorded, untagged animals are given tags, and the entire sample is returned to the population. If the population is being exploited then additional recapture records may be available in the form of tag returns from the harvest taken in the interims between t_i and t_{i+1}. Excluding these additional records, the survival rate between t_i and t_{i+1} is estimated by $\hat{s}_i = \hat{N}_{i+1}/\hat{N}_i$, where $\hat{N}_i$ is the maximum likelihood estimator of population size at time t_i. The sampling distribution of $\hat{s}_i$ is examined here and shown empirically to be closely approximated by a gamma density function having the same first two moments. An estimator $\hat{s}_i^*$ utilizing reported recaptures from the harvest is derived heuristically. In this preliminary report the results are present in outline form only.

Statistical Methods for a Tag-Recapture Experiment on a Population Closed to Recruitment: Preliminary Report

D. S. Robson

BU-374-M June, 1971

INTRODUCTION

Survival rates in an animal population or subpopulation which is closed to recruitment but open to mortality may be estimated from a k-sample tag-recapture experiment. At time t_i a batch of m_i animals are captured, the tag numbers of all recaptures are recorded, untagged animals are given tags, and the entire sample is returned to the population. If the population is being exploited then additional recapture records may be available in the form of tag returns from the harvest taken in the interims between t_i and t_{i+1}. Excluding these additional records, the survival rate between t_i and t_{i+1} is estimated by $\hat{s}_i = \hat{N}_{i+1}/\hat{N}_i$, where $\hat{N}_i$ is the maximum likelihood estimator of population size at time t_i. The sampling distribution of $\hat{s}_i$ is examined here and shown empirically to be closely approximated by a gamma density function having the same first two moments. An estimator $\hat{s}_i^*$ utilizing reported recaptures from the harvest is derived heuristically. In this preliminary report the results are present in outline form only.

TABLE 1. NOTATION for the multiple sample tag-recapture model

Time of sampling	Population size	Survival rate	Sample size	Number of fish which are subsequently recaptured
t_1	N_1		m_1	R_1
		$s_1 = \frac{N_2}{N_1}$		
t_2	N_2		m_2	R_2
		$s_2 = \frac{N_3}{N_2}$		
t_3	N_3		m_3	R_3
.	.	$s_3 = \frac{N_4}{N_3}$	.	.
.	.	.	.	.
.	.	.	.	.
		.		
		.		
t_{k-1}	N_{k-1}		m_{k-1}	R_{k-1}
		$s_{k-1} = \frac{N_k}{N_{k-1}}$		
t_k	N_k		m_k	$R_k = 0$

$$C_i = (m_{i+1} - R_{i+1}) + (m_{i+2} - R_{i+2}) + \dots + (m_k - R_k)$$

= number of (distinct) fish captured <u>after</u> the i'th sample

ALTERNATE NOTATION for unions and intersections of samples

General: An underlined letter $\underline{M}$ denotes a set of $M = \phi(\underline{M})$ members.

Particular: $\underline{m}_1$ is the set of $m_1 = \phi(\underline{m}_1)$ fish collected in the first sample.

$\underline{C}_1 = \underline{m}_2 \cup \underline{m}_3 \cup \dots \cup \underline{m}_k$ is the set of $C_1 = \phi(\underline{C}_1)$ fish captured after the first sample.

$\underline{R}_1 = \underline{m}_1 \cap \underline{C}_1$ is the set of $R_1 = \phi(\underline{R}_1)$ fish captured in the first sample and subsequently recaptured.

TABLE 2. NOTATION for ANGLER TAG RETURNS

Time of sampling	Number of angler tag recoveries during interval between samples	Number of angler tag recoveries made in subsequent intervals
t_1	$a_1 = \phi(\underline{a}_1)$	$A_1 = \phi(\underline{a}_1 \cup \underline{a}_2 \cup \dots \cup \underline{a}_k) = \sum_{i=1}^{k} a_i$
t_2	$a_2 = \phi(\underline{a}_2)$	$A_2 = \phi(\underline{a}_2 \cup \underline{a}_3 \cup \dots \cup \underline{a}_k) = \sum_{i=2}^{k}$
t_3	$a_3 = \phi(\underline{a}_3)$	$A_3 = \phi(\underline{a}_3 \cup \underline{a}_4 \cup \dots \cup \underline{a}_k) = \sum_{i=3}^{k} a_i$
$\vdots$	$\vdots$	$\vdots$
t_{k-1}	$a_{k-1} = \phi(\underline{a}_{k-1})$	$A_{k-1} = \phi(\underline{a}_{k-1} \cup \underline{a}_k)$
t_k	$a_k = \phi(\underline{a}_k)$	$A_k = \phi(\underline{a}_k) = \underline{a}_k$

ASSUMPTIONS concerning the multiple sample tag-recapture model:

1. No recruitment into the population being sampled.

2. Each sample is a random sample from the population existing at that time.

3. Capture, tag, and release do not affect the chances for survival of a fish.

4. Fish do not lose their tags.

ASSUMPTIONS concerning angler tag returns:

1. Time of capture is reported to the correct interval between samples.

2. When a tagged fish is removed by angling the chances that the tag will be reported are independent of the time of tagging.

TABLE 3. Illustration of NOTATION for RECAPTURE HISTORIES, k = 4 samples (1 stands for "captured", 0 for "not captured")

Capture history Sample 1	Sample 2	Sample 3	Sample 4	Observed number of cases	Expected number of cases
1	0	0	0	X_{1000}	m_1-R_1
0	1	0	0	X_{0100}	$(m_2-R_2)\,\frac{C_1-R_1}{C_1}$
1	1	0	0	X_{1100}	$(m_2-R_2)\,\frac{R_1}{C_1}$
0	0	1	0	X_{0010}	$(m_3-R_3)\,\frac{(C_1-R_1)(C_2-R_2)}{C_1C_2}$
1	0	1	0	X_{1010}	$(m_3-R_3)\,\frac{R_1(C_2-R_2)}{C_1C_2}$
0	1	1	0	X_{0110}	$(m_3-R_3)\,\frac{(C_1-R_1)R_2}{C_1C_2}$
1	1	1	0	X_{1110}	$(m_3-R_3)\,\frac{R_1R_2}{C_1C_2}$
0	0	0	1	X_{0001}	$(m_4-R_4)\,\frac{(C_1-R_1)(C_2-R_2)(C_3-R_3)}{C_1C_2C_3}$
1	0	0	1	X_{1001}	$(m_4-R_4)\,\frac{R_1(C_2-R_2)(C_3-R_3)}{C_1C_2C_3}$
0	1	0	1	X_{0101}	$(m_4-R_4)\,\frac{(C_1-R_1)R_2(C_3-R_3)}{C_1C_2C_3}$
1	1	0	1	X_{1101}	$(m_4-R_4)\,\frac{R_1R_2(C_3-R_3)}{C_1C_2C_3}$
0	0	1	1	X_{0011}	$(m_4-R_4)\,\frac{(C_1-R_1)(C_2-R_2)R_3}{C_1C_2C_3}$
1	0	1	1	X_{1011}	$(m_4-R_4)\,\frac{R_1(C_2-R_2)R_3}{C_1C_2C_3}$
0	1	1	1	X_{0111}	$(m_4-R_4)\,\frac{(C_1-R_1)R_2R_3}{C_1C_2C_3}$
1	1	1	1	X_{1111}	$(m_4-R_4)\,\frac{R_1R_2R_3}{C_1C_2C_3}$

CHI-SQUARE goodness of fit test:

$$\chi^2 = \sum \frac{(\text{Observed} - \text{Expected})^2}{\text{Expected}}$$

$$= \sum_{i=2}^{k} \sum_{\delta_1,\ldots,\delta_{i-1}} \frac{\left[X_{\delta_1\ldots\delta_{i-1}10\ldots0} - (m_i - R_i)\prod_{j=1}^{i-1} \frac{R_j^{\delta_j}(C_j - R_j)^{1-\delta_j}}{C_j}\right]^2}{(m_i - R_i)\prod_{j=1}^{i-1} \frac{R_j^{\delta_j}(C_j - R_j)^{1-\delta_j}}{C_j}}$$

Degrees of freedom = $2^k - 2k$

<u>NOTE</u>:

$$\sum_{\delta_1,\ldots,\delta_{i-1},\delta_{i+1},\ldots,\delta_k} X_{\delta_1\ldots\delta_{i-1}1\delta_{i+1}\ldots\delta_k} = m_i \quad \text{for } i = 1,\ldots,k$$

$$\sum_{\delta_1,\ldots,\delta_{i-1}} X_{\delta_1\ldots\delta_{i-1}10\ldots0} = m_i - R_i \qquad \text{for } i = 1,\ldots,k$$

and

$$P[\{X_{\delta_1\ldots\delta_k}\} = \{x_{\delta_1\ldots\delta_k}\} | \{m_i\},\{R_i\}]$$

$$= \prod_{i=1}^{k-1} \frac{(m_{i+1} - R_{i+1})}{\binom{C_i}{R_i}} \Bigg/ \prod_{\delta_1,\ldots,\delta_i} (x_{\delta_1\ldots\delta_i\,10\ldots0})!$$

ESTIMATION

Maximum likelihood estimators:

$$\hat{N}_i = \frac{m_i C_i}{R_i} \qquad \hat{s}_i = \frac{\hat{N}_{i+1}}{\hat{N}_i}$$

Percentage bias in maximum likelihood estimators:

$$\text{Bias in } \hat{N}_i \doteqdot 100 \exp\left(-\frac{m_i C_i}{N_i}\right)\% \qquad \text{Bias in } \hat{s}_i \doteqdot 100 \exp\left(\frac{m_{i+1} C_{i+1}}{N_{i+1}}\right)\%$$

Adjustments for bias:

$$\tilde{N}_i = \frac{(m_i + 1)(C_i + 1)}{R_i + 1} - 1 \qquad \tilde{s}_i = \frac{\tilde{N}_{i+1}}{\hat{N}_i}$$

$$\text{Bias in } \tilde{N}_i = \text{Bias in } \tilde{s}_{i-1}$$

$$= 100 \frac{N_i - C_i}{N_i} \frac{\binom{N_i - C_i - 1}{m_i}}{\binom{N_i}{m_i}} \%$$

$$= 100 \frac{N_i - m_i}{N_i} \frac{\binom{N_i - m_i - 1}{C_i}}{\binom{N_i}{C_i}}$$

$$= 0 \text{ if } m_i + C_i \geq N_i$$

$$\doteqdot 100 \exp\left(-\frac{m_i C_i}{N_i}\right) \text{ if } m_i + C_i \ll N_i$$

Variance estimators:

$$\hat{\sigma}^2_{\tilde{N}_i} = (\tilde{N}_i + 1)(\tilde{N}_i + 2) - \frac{(m_i + 1)(m_i + 2)(C_i + 1)(C_i + 2)}{(R_i + 1)(R_i + 2)}$$

(Bias in $\hat{\sigma}^2_{\tilde{N}_i}$ is zero if $m_i + C_i \geq N_i$)

$$\hat{\sigma}^2_{\tilde{s}_i} = \frac{R_i(R_i - 1)}{m_i(m_i - 1)C_i(C_i - 1)} \left[\hat{\sigma}^2_{\tilde{N}_{i+1}} - \frac{m_i + C_i}{m_i C_i} \tilde{N}^2_{i+1} \right]$$

$$+ \frac{R_i}{m_i C_i}\left[\frac{1}{m_i C_i} + \frac{(R_i - 1)(R_i - 2)}{(m_i - 1)(m_i - 2)(C_i - 1)(C_i - 2)}\right]\tilde{N}^2_{i+1}$$

Interval estimator of survival rate:

The sampling distribution of $2s_i\tilde{s}_i/\sigma^2_{\tilde{s}_i}$ is closely approximated by a gamma distribution

$$f(y) = \frac{y^{\lambda-1}}{2^{\lambda}\Gamma(\lambda)}\exp(-\frac{y}{2}) \qquad , \quad \lambda = s_i^2/\sigma^2_{\tilde{s}_i}$$

which, for integer values of λ, is the chi-square distribution on 2λ degrees of freedom. Hence, with $\hat{\lambda} = [\tilde{s}_i^2/\hat{\sigma}^2_{\tilde{s}_i}]$ a 95% confidence interval estimator of s_i is given by

$$\frac{2\hat{\lambda}\tilde{s}_i}{\chi^2_{.975}(2\hat{\lambda}\text{ d.f.})} < s_i < \frac{2\hat{\lambda}\tilde{s}_i}{\chi^2_{.025}(2\hat{\lambda}\text{ d.f.})}$$

If $\hat{\lambda}$ is large (say $\hat{\lambda} > 50$) this becomes essentially the normal approximation

$$\tilde{s}_i - 1.96\,\hat{\sigma}_{\tilde{s}_i} < s_i < \tilde{s}_i + 1.96\,\hat{\sigma}_{\tilde{s}_i}$$

Estimation of survival rate utilizing ANGLER TAG RETURNS:

$$\hat{s}_i^* = \frac{\phi(\underline{m}_i \cap \underline{m}_{i+1})\phi[\underline{m}_i \cap (\underline{C}_{i+1} \cup \underline{A}_{i+1})]}{m_i\,\phi[(\underline{m}_i \cap \underline{m}_{i+1})\cap(\underline{C}_{i+1} \cup \underline{A}_{i+1})]}$$

or, adjusted for bias,

$$\tilde{s}_i^* = \frac{1}{m_i}\left[\frac{\{\phi(\underline{m}_i \cap \underline{m}_{i+1}) + 1\}\{\phi[\underline{m}_i \cap (\underline{C}_{i+1} \cup \underline{A}_{i+1})] + 1\}}{1 + \phi[(\underline{m}_i \cap \underline{m}_{i+1})\cap(\underline{C}_{i+1} \cup \underline{A}_{i+1})]} - 1\right]$$

The latter is an exactly unbiased estimator of the proportion of the m_i fish released at time t_i that survive to time t_{i+1}, provided that the number surviving is less than or equal to

$$\phi(\underline{m}_i \cap \underline{m}_{i+1}) + \phi[\underline{m}_i \cap (\underline{C}_{i+1} \cup \underline{A}_{i+1})]$$

Variance estimation utilizing ANGLER TAG RETURNS:

In the notation

$$\tilde{s}_i^* = \frac{1}{m_i}\left[\frac{(u_i + 1)(v_i + 1)}{w_i + 1} - 1\right] = \frac{1}{m_i}\left[q_i - 1\right]$$

we get

$$\hat{\sigma}^2_{\tilde{s}_i^*} = \frac{1}{m_i^2}\left[q_i(q_i + 1) - \frac{(u_i + 1)(u_i + 2)(v_i + 1)(v_i + 2)}{(w_i + 1)(w_i + 2)}\right]$$

which is unbiased if $\tilde{s}_i^*$ is unbiased.

A RANK-SUM TEST OF THE EFFECT OF HOLDING TIME

If the m_1 fish tagged and released at time t_1 are held in a tank and released individually as they are tagged then holding time increases directly with the serial tag numbers. If holding time does effect survival then the lower tag numbers should predominate in the set $R_1 \cup (\underline{A}_1 \cap \underline{m}_1)$ of recaptures. Partitioning $\underline{m}_1$ into $\underline{R}_1 \cup (\underline{A}_1 \cap \underline{m}_1)$ and its complement then provides the framework for Wilcoxon's (1964) two-sample rank-sum test on the rank-order with respect to holding time.

A RANK-SUM TEST OF RECRUITMENT INTO THE CATCHABLE SIZE CLASS

If the m_i fish captured at time t_i are ranked according to body size and then partitioned into two samples, tagged and untagged, then the rank-sum for either sample may be tested against Wilcoxon's tables to determine whether the newly captured fish are predominantly smaller than the recaptures.

AN EXCEEDANCE TEST FOR RECRUITMENT TO CATCHABLE SIZE

If recruitment through growth to catchable size (particularly in the youngest age group being sampled)does occur then in the sample taken at time t_i the number of untagged fish ranging in size between successively larger tagged fish should be decreasing. In contrast, with no such recruitment this number should be constant. A test for an excessive number of untagged fish in the lower tail of the size distribution of tagged fish is provided by Gumbel and von Schelling's "theory of exceedances" (1950) and is described in the context of the present problem by Robson and Flick (1965).

RANK-SUM TEST OF THE EFFECT OF FISH SIZE ON PROBABILITY OF RECAPTURE

If the m_i fish taken in the i'th sample are ranked according to body size and then partitioned into two samples according to whether or not a fish is subsequently recaptured, then Wilcoxon's rank-sum procedure may be applied to test whether these two samples represent the same size distribution.

TEST FOR TAGGING MORTALITY IN A 2 X 2 TABLE

The m_2 fish taken in the second sample may be partitioned into a 2 X 2 array when classified according to their capture history in the first and third samples,

		First sample: captured	First sample: not captured
Third sample	captured	X_{111}	X_{011}
	not captured	X_{110}	X_{010}

and tested by contingency chi-square or, if the numbers are small, by the exact test for 2 X 2 tables. Tagging mortality (over and above mortality caused by handling fish already tagged) would produce the relation

$$\frac{X_{111}}{X_{110} + X_{111}} - \frac{X_{011}}{X_{011} + X_{010}} > 0$$

though other factors such as capture proneness could also produce a significant difference.

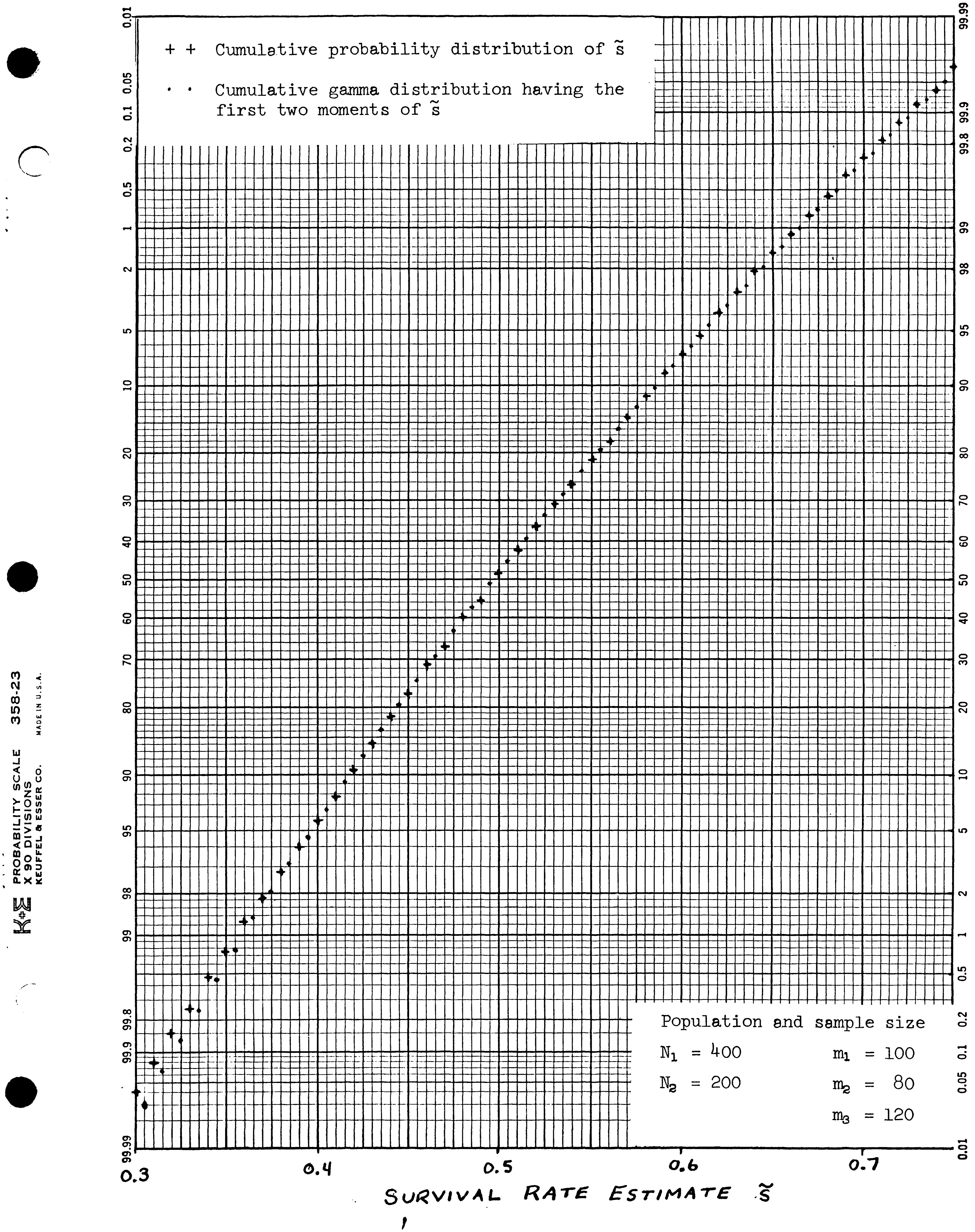
References

Jolly, G. M., "Explicit maximum likelihood estimates from capture-recapture data with both death and immigration - stochastic model". Mimeographed draft. Agricultural Research Council Unit of Statistics, Aberdeen, 1964.

Wilcoxon, F. and Wilcox, R. A., Some rapid approximate statistical procedures. Lederle Laboratories, Pearl River, N.Y., revised 1964.

Gumbel, E. J. and H. von Schelling, "The distribution of the number of exceedances", Annals of Mathematical Statistics 21, p. 247-262, 1950.

Robson, D. S. and W. A. Flick, "A nonparametric statistical method for culling recruits from a mark-recapture experiment". Mimeographed draft. Biometrics Unit, Cornell University, 1964.



+ + Cumulative probability distribution of $\tilde{s}$
· · Cumulative gamma distribution having the first two moments of $\tilde{s}$
Population and sample size
N_1 = 400
N_2 = 200
m_1 = 100
m_2 = 80
m_3 = 120
0.3
0.4
0.5
0.6
0.7
SURVIVAL RATE ESTIMATE $\tilde{s}$
K&E PROBABILITY SCALE X 90 DIVISIONS 358-23
KEUFFEL & ESSER CO. MADE IN U.S.A.